

# International R&D spillovers from trade

Esther Ann Nisja Bøler



MASTER THESIS FOR THE DEGREE  
MASTER OF PHILOSOPHY IN ECONOMICS

Department of Economics

University of Oslo

August 2010

# Preface

This thesis is written as part of the project *R&D, Industry Dynamics and Public Policy* (6512) at the Ragnar Frisch Centre for Economic Research. The project is funded by the Norwegian Research Council, and is a collaboration between the Department of Economics at the University of Oslo, the BI Norwegian School of Management and the Ragnar Frisch Centre for Economic Research.

I would like to thank my supervisor Karen Helene Ulltveit-Moe for involving me in the project, and for fruitful cooperation throughout the process. I also wish to thank my co-supervisor Andreas Moxnes, for the estimations of the TFP variable and for always having the time to answer my questions. I am grateful to everyone at the The Frisch Centre for welcoming me into the work environment. I would also like to thank the Centre for Equality, Social Organization and Performance (ESOP) at the Department of Economics at the University of Oslo, for being a great employer during my master studies. I have been given a lot of interesting and challenging assignments that have proved useful while working with the thesis.

## Summary

In an influential paper from 1995, “International R&D spillovers”, David Coe and Elhanan Helpman use data on trade and R&D expenditures for 22 countries to estimate the effects of a country’s own R&D effort and the R&D effort of its trade partners on the country’s TFP. They use a model of endogenous, innovation-driven growth as basis for the empirical equations. The main idea is that a country benefits from the cumulative R&D stock of its trade partners as well as its own R&D stock, through spillover effects from trade in intermediate inputs.

The intermediate inputs can be horizontally differentiated, which means that a new input is equally good as an old, and investment in R&D increases the number of available varieties. The inputs can also be vertically differentiated, so that the development of a new input will replace an old one because of its increased quality. In the endogenous growth framework, R&D investments are made by firms seeking monopoly profits in imperfectly competitive markets. This R&D effort generates a product or process than can be patented and give rise to a profit for the firm, as well as a non-appropriable product that will increase the country’s stock of knowledge. This non-appropriable part is what generates the spillover effect. In the case of horizontally differentiated inputs, the investments in R&D increase the stock of knowledge through an increase in the number of available inputs, which in turn decreases future R&D costs. With vertically differentiated inputs, the development of a higher quality input will enable future entrepreneurs to build on a higher quality foundation. In both cases, the stock of knowledge increases and there are spillovers from current to future R&D activities.

Coe and Helpman (1995) argue that since intermediate inputs are traded internationally, a country’s productivity depends on its own R&D stock as well as the R&D stock of its trade partners. The domestic R&D stock is constructed as the sum of the country’s cumulative R&D expenditures, multiplied by a constant depreciation factor. In the basic specification used in Coe and Helpman (1995), the foreign R&D stock variable is constructed as a weighted sum of the R&D stocks of the country’s trade partners, where the weight is the imports from one country as share of the total imports. They find a positive and significant effect of both R&D stocks on the country’s TFP, and interpret this as evidence that there are significant international R&D spillovers from trade.

Norway is one of the countries used in the paper. I use panel data on Norwegian firms to see whether this effect of imported R&D spillovers can be found on the firm level in Norway. The data set consists of data from the accounting and manufacturing statistics, as well as data on trade and R&D expenditures for Norwegian firms. This is combined with data on R&D expenditures on the sector level for 22 member countries of the OECD. The variables are constructed on the basis of the variables used in Coe and Helpman (1995), with some adjustments to fit the data set at hand.

I estimate the basic specification from Coe and Helpman (1995), using the fixed effects estimator. The software used for the estimations is Stata 11. I find a positive and significant effect of the firm's own R&D stock. However, the foreign R&D stock appears to have no significant impact on the productivity of the firms. This indicates that any spillover effects that Norway might get from foreign R&D, are not transmitted through each firm's imported intermediate inputs.

I also estimate several extensions of the original specification. In the first extension, I substitute the firm's own R&D stock with the lagged R&D expenditure, because of the nonstationarity of the R&D stock variable. In the second, the weights are changed to reflect the fact that a firm can purchase intermediate inputs domestically as well as internationally. The foreign R&D stock variable is then constructed using the imports from one country as share of total purchase of intermediate inputs as weights. The third extension is aggregating the foreign R&D stock variable to the sector level, to see if I can detect any intra-industry spillovers that are transmitted through trade in intermediate inputs.

The positive and significant effect of the firm's own R&D stock is stable in the various specifications. The foreign R&D stock variable is however not statistically significant at any of the conventional levels. This confirms the indication that any spillover effects do not emanate from each firm's imports of intermediate inputs, and it also indicates that any spillover effects have to appear at a more aggregated level.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Theoretical Framework</b>	<b>4</b>
2.1	Theory . . . . .	4
2.2	Theoretical Model . . . . .	7
2.2.1	Basic Model . . . . .	8
2.2.2	Coe and Helpman's Specification . . . . .	12
<b>3</b>	<b>The Data</b>	<b>16</b>
3.1	Description of the Data Sets . . . . .	16
3.2	Sample Selection . . . . .	18
3.3	The Variables . . . . .	20
3.3.1	TFP . . . . .	20
3.3.2	R&D stock . . . . .	21
3.3.3	Foreign R&D stock . . . . .	22
<b>4</b>	<b>Empirical Strategy and Results</b>	<b>23</b>
4.1	Clustered Standard Errors . . . . .	26
4.2	Extension I: Nonstationarity . . . . .	29
4.3	Extension II: Different weights . . . . .	33
4.4	Extension III: Aggregating the foreign R&D stock variable . . . . .	34
4.5	Remarks . . . . .	36
<b>5</b>	<b>Conclusion</b>	<b>38</b>

# 1 Introduction

In an influential paper from 1995, “International R&D spillovers”, David Coe and Elhanan Helpman use data on trade and research and development (R&D) expenditures for 22 countries to estimate the effects of a country’s own R&D effort and the R&D effort of its trade partners on the country’s total factor productivity (TFP). They use a model of endogenous, innovation-driven growth as basis for the empirical equations. The main idea is that a country benefits from the cumulative R&D stock of its trade partners as well as its own R&D stock, through spillover effects from trade in intermediate inputs.

The intermediate inputs can be horizontally differentiated, which means that a new input is equally good as an old, and investment in R&D increases the number of available varieties. The inputs can also be vertically differentiated, so that the development of a new input will replace an old one because of its increased quality. In the endogenous growth framework, R&D investments are made by firms seeking monopoly profits in imperfectly competitive markets. This R&D effort generates a product or process than can be patented and give rise to a profit for the firm, as well as a non-appropriable product that will increase the country’s stock of knowledge. This non-appropriable part is what generates the spillover effect. In the case of horizontally differentiated inputs, the investments in R&D increase the stock of knowledge through an increase in the number of available inputs, which in turn decreases future R&D costs. With vertically differentiated inputs, the development of a higher quality input will enable future entrepreneurs to build on a higher quality foundation. In both cases, the stock of knowledge increases and there are spillovers from current to future R&D activities.

Coe and Helpman (1995) argue that since intermediate inputs are traded internationally, a country’s productivity depends on its own R&D stock as well as the R&D stock of its trade partners. The domestic R&D stock is constructed as the sum of the country’s cumulative R&D expenditures, multiplied by a constant depreciation factor. In the basic specification used in Coe and Helpman (1995), the foreign R&D stock variable is constructed as a weighted sum of the R&D stocks of the country’s trade partners, where the weight is the imports from one country as share of the total imports. They find a positive and significant effect of both R&D stocks on the country’s TFP, and interpret this as evidence that there are significant international

R&D spillovers from trade.

Norway is one of the countries used in the paper. I use panel data on Norwegian firms to see whether this effect of imported R&D spillovers can be found on the firm level in Norway. The data set consists of data from the accounting and manufacturing statistics, as well as data on trade and R&D expenditures for Norwegian firms. This is combined with data on R&D expenditures on the sector level for 22 member countries of the Organization for Economic Co-operation and Development (OECD). The variables are constructed on the basis of the variables used in Coe and Helpman (1995), with some adjustments to fit the data set at hand.

I estimate the basic specification from Coe and Helpman (1995), using the fixed effects estimator. The software used for the estimations is Stata 11. I find a positive and significant effect of the firm's own R&D stock. However, the foreign R&D stock appears to have no significant impact on the productivity of the firms. This indicates that any spillover effects that Norway might get from foreign R&D, are not transmitted through each firm's imported intermediate inputs.

I also estimate several extensions of the original specification. In the first extension, I substitute the firm's own R&D stock with the lagged R&D expenditure, because of the nonstationarity of the R&D stock variable. In the second, the weights are changed to reflect the fact that a firm can purchase intermediate inputs domestically as well as internationally. The foreign R&D stock variable is then constructed using the imports from one country as share of total purchase of intermediate inputs as weights. The third extension is aggregating the foreign R&D stock variable to the sector level, to see if I can detect any intra-industry spillovers that are transmitted through trade in intermediate inputs.

The positive and significant effect of the firm's own R&D stock is stable in the various specifications. The foreign R&D stock variable is however not statistically significant at any of the conventional levels. This confirms the indication that any spillover effects do not emanate from each firm's imports of intermediate inputs, and it also indicates that any spillover effects have to appear at a more aggregated level.

The thesis has the following structure: The theoretical background, the endogenous growth model and the specification used in Coe and Helpman (1995) is explained

in further detail in Section 2. The data set and variables used in the regressions is described in Section 3. Section 4 consists of the empirical strategy, equations and results, as well as the several extensions of the basic specification. Section 5 concludes.



## 2 Theoretical Framework

### 2.1 Theory

In neoclassical growth theory, technological change is taken as exogenous. In the Solow framework, for instance, the long-run growth rate is determined by the investment in and the depreciation of capital, assuming a rate of technological growth. The underlying source of this technological change is not incorporated in the models. In contrast to this, the endogenous growth theory endogenizes technological change and treats growth as “driven by technological change that arises from intentional investment decisions made by profit-maximizing agents” (Romer, 1990). Among the most influential theoretical work are Romer (1986, 1990), Lucas (1988) and Aghion and Howitt (1992), while “Innovation and growth in the global economy” by Grossman and Helpman (1991) is a seminal work on endogenous growth in an open economy.

In this framework, R&D investments are thought to affect productivity through two channels. The first is by stimulating innovation, of for instance new products or production processes. The second is by facilitating the assimilation of other firms’ or countries’ inventions, the so-called *absorptive capacity*. These are often called the two “faces” of R&D.<sup>1</sup>

The R&D investments of a firm or country is likely to affect its own productivity, and it can also influence the productivity of other firms or countries, through so-called *spillover effects*. A spillover effect is a positive externality, which means that the actions of one agent positively influences another agent, without this effect being explicitly included in any of the agents’ decisions. So what exactly constitute these spillover effects from R&D investments? In the models of Romer (1986) and Lucas (1988), aggregate production exhibits increasing returns to scale – doubling the use of all inputs will more than double the output. This implies that firms will seek to grow infinitely large, and there will not be perfect competition in the market, as larger firms always will be more productive. To avoid this, it is assumed that the returns to scale are external to each firm. In that way, there can be perfect competition between the firms, but increasing returns to scale for the aggregate R&D

---

<sup>1</sup>See for instance Cohen and Levinthal (1989) and Griffith et al. (2004) for empirical studies trying to separate the two effects.

investments. This is the source of spillovers in these models.

However, perfect competition is not an attractive feature when trying to model commercial R&D investments, even though allowing for imperfect competition complicates the models significantly. In Helpman's words: "Such complications are unavoidable, however, when we wish to deal explicitly with profit seeking investment in innovation, as we should, given the rising importance of commercial R&D in the industrial world" (Helpman, 1992).

In the endogenous growth framework described in Section 2.2, R&D investments are made by firms seeking monopoly profits in imperfectly competitive markets. This R&D effort is thought to generate a product or process that can be patented and give rise to a profit for the firm, as well as a non-appropriable product that will increase the country's stock of knowledge. A new technology can be patented, but a new way of using an already developed material cannot. This non-appropriable part is what generates the spillover effect in this framework. It also makes the private return to R&D different from the social return, which includes all these spillover effects.

Coe and Helpman write the following about the relationship between R&D and productivity: "Own R&D produces traded and nontraded goods and services that bring about more effective use of existing resources and thereby raises a country's productivity level. In addition, own R&D enhances a country's benefits from foreign technical advances, and the better a country takes advantage of technological advances in the rest of the world the more productive it becomes. The benefits from foreign R&D can be both direct and indirect. Direct benefits consist of learning about new technologies and materials, production processes, or organizational methods. Indirect benefits emanate from imports of goods and services that have been developed by trade partners. In either case foreign R&D affects a country's productivity" (Coe and Helpman, 1995). In their paper, Coe and Helpman concentrate on this indirect effect of foreign R&D when they focus on the trade in intermediate inputs as the transmission mechanism for spillovers.

According to Griliches (1992), there are two kinds of "spillovers" that are often confused. One type occurs if an industry buys inputs that are priced lower than their quality dictates. In that way, the industry's productivity will increase without

the correct share of the increase being attributed to the inputs. However, Griliches argues that these so-called *rent spillovers* are simply consequences of measurement problems. They arise because the prices have not adjusted to fully reflect the quality improvements of the inputs. The proper spillover effects, *knowledge spillovers*, are generated when researchers in different industries borrow ideas from each other. They can be caused by for instance poor patent protection, the inability to keep innovations secret or reverse engineering (Cincera and van Pottelsberghe de la Potterie, 2001).

There has been a lot of empirical work trying to determine the scope and magnitude of domestic and international spillover effects of R&D.<sup>2</sup> Domestic spillovers exist when a firm or sector in a country benefits from the R&D investments of other firms or sectors in the same country. International R&D spillovers occur when the R&D investment in one country affects the productivity of another country. In the literature, these spillovers are investigated using a number of different approaches. Trade, foreign direct investment (FDI), international R&D collaborations, patent sales, licencing agreements, joint ventures and migration of researchers are some of the mechanisms thought to transfer the spillovers. Coe and Helpman (1995) and Coe et al. (1997) are among those who argue that the spillovers are transmitted through trade, Jaffe et al. (1993) and Eaton and Kortum (1996) focus on patenting, Lichtenberg and van Pottelsberghe de la Potterie (1996) concentrate on FDI and trade, while Bernstein and Mohnen (1998) investigate the spillovers between Japanese and U.S. R&D intensive sectors using a general specification that does not rely on any specific mechanism of spillover transfer.

The Coe and Helpman paper has been an important contribution to this literature. It has inspired a lot of later research, and it has also received a fair bit of criticism. The use of trade flows as the transmission channel for spillovers from R&D has been particularly debated. The main objection is that the trade patterns in this case are endogenous.<sup>3</sup> The country will choose to import from countries performing a lot of R&D because of their bigger range, or better quality, of intermediate inputs. This will lead to biased estimates of the spillover effects of R&D. Among the critics is Keller (1998), who uses randomly generated patterns of trade and estimates

---

<sup>2</sup>See Griliches (1992), Nadiri (1993) and Cincera and van Pottelsberghe de la Potterie (2001) for surveys.

<sup>3</sup>See for instance Bernstein and Mohnen (1998) and Keller (1998).

the same relationship as Coe and Helpman. He finds an even stronger effect of imported R&D than the effect found in Coe and Helpman (1995), which is an indication that international trade patterns are not as important in the transmission of R&D spillovers as Coe and Helpman conclude.

In a follow-up paper from 2009, Coe, Helpman and Hoffmaister address this issue. They refer to Coe and Hoffmaister (1999), who show that the weights used by Keller are not really random, and that there is no significant effect of foreign R&D when using proper randomly generated weights. They also argue that their use of so-called *cointegration* techniques produces superconsistent estimates which are robust to endogeneity. This method will be discussed in more detail in Section 4.2. My data set does not enable the use of cointegration. I will however argue that this is not as big of a problem in my specification, when using firm TFP as the dependent variable. The firm decides its foreign trade partners based on the relevant firms' products, prices, and so forth, not based on the R&D performance of the foreign country or sector as a whole.

Several studies have concluded that the absorptive capacity mentioned above is important for a country to be able to benefit from the international spillovers from R&D.<sup>4</sup> In other words, for a country's productivity to be enhanced by the investments of its trade partners, it is vital that the country itself invests in R&D. Otherwise it will not acquire the tacit knowledge needed to take advantage of the spillover effects from the imported R&D (Griffith et al., 2004). Based on this, I would not suspect to find much spillover effects from imported R&D in Norway if Norway itself did not invest in R&D. According to data from OECD, this is not a concern. Norway invests more in R&D than the European Union average, when looking at both gross domestic expenditure on R&D per capita, and total R&D personnel in relation to total employment.<sup>5</sup>

## 2.2 Theoretical Model

The theoretical specification used in Coe and Helpman's paper builds on a model of innovation-driven growth. I will give a brief description of the model in this chapter. For a more detailed exposition, see Grossman and Helpman (1991) and Helpman

---

<sup>4</sup>See for instance Jaffe (1986) and Griffith et al. (2004).

<sup>5</sup>See Appendix A for tables.

(1992).<sup>6</sup>

### 2.2.1 Basic Model

The model builds on a Dixit-Stiglitz index of differentiated products:

$$D = \left[ \int_0^n x(j)^\alpha dj \right]^{1/\alpha}, \quad 0 < \alpha < 1 \quad (1)$$

where  $x(j)$  is the quantity of variety  $j$ ,  $n$  is the number of available brands, and  $\alpha$  is a parameter that measures to what degree the varieties can substitute each other. This index implies a constant elasticity of substitution (CES) between any two brands equal to  $\varepsilon = 1/(1 - \alpha) > 1$ .<sup>7</sup> That this elasticity of substitution is larger than one, means that the varieties substitute well for each other.

This index yields CES demand functions with marginal revenue  $MR(j) = \alpha p(j)$ , where  $p(j)$  denotes the price of variety  $j$ . The firms in this model, besides manufacturing the already developed products, can also develop blueprints for new varieties. They only use one primary factor of production, labor, and the labor can be employed either in production or in R&D activities. Helpman (1992) normalizes the labor input-output coefficient in manufacturing to one, so that all varieties are produced using one unit of labor per unit of output. Then marginal cost per unit of output  $MC(j)$  equals the wage rate  $w$ .

Investment in R&D leads to development of new varieties, and once a firm has developed a new variety, it gets monopoly power in the supply of this brand. The monopolistic competition implies that in equilibrium, when each firm maximizes its profits, marginal cost equals marginal return. This implies that all varieties are priced equally at

$$p = w/\alpha. \quad (2)$$

The profits for a firm is then given by

$$\pi = (1 - \alpha) \frac{pX}{n}, \quad (3)$$

where  $X$  is the aggregate output of products. This expression shows that as the number of available varieties grows, the profit for each firm is reduced.

---

<sup>6</sup>Coe and Helpman also refer to these sources, and do not provide anything but a short description in their paper. My description of the model will mostly rely on Helpman (1992).

<sup>7</sup>See Appendix B for the calculations.

The firm's value is given by the present value of its profits:

$$v(t) = \int_t^\infty e^{R(\tau,t)} \pi(\tau) d\tau, \quad (4)$$

where  $R(\tau, t) = \int_t^\tau r(z) dz$  is the discount rate from time  $\tau$  to  $t$ . Differentiation of this expression with respect to  $t$  yields:

$$\frac{\pi}{v} + \frac{\dot{v}}{v} = r. \quad (5)$$

This is the no-arbitrage condition, which ensures that the rate of return to the ownership of a firm equals the nominal interest rate  $r$ .

The gain from investing in R&D has to exceed the cost of doing so for any agent to be willing to engage in R&D activities. If an entrepreneur uses  $l$  units of labor in R&D for a period of  $dt$ , the cost of this will be  $wl dt$ . Assume that the productivity of the labor engaged in R&D rises with the stock of knowledge available,  $K_n$ . Then the total gain is  $vl(K_n/a)dt$ , where  $K_n/a$  is the productivity of the labor in the R&D activity. This implies that  $l$  will be as high as possible whenever  $vK_n/a > w$ , and zero when  $vK_n/a < w$ . In equilibrium, we have to have

$$\frac{wa}{K_n} \geq v, \quad (6)$$

which has to be satisfied with equality for there to be any investment in R&D.

Labor market clearing implies that the employment in R&D and manufacturing equals the total labor supply:

$$\frac{a}{K_n} \dot{n} + X = L, \quad (7)$$

where  $a/K_n$  is the labor requirement to produce one unit of output in the R&D activities, which has to be multiplied by the number of blueprints created to get the total labor demand in R&D.  $X$ , the aggregate output of products, equals the labor employed in manufacturing since the input-output coefficient equals one, and  $L$  is the total supply of labor.

In this endogenous growth model, savings are invested in R&D. The consumers have the following intertemporal utility function:

$$U_t = \int_t^\infty e^{-\rho(\tau-t)} u(\tau) d\tau, \quad (8)$$

where  $\rho$  is the subjective discount rate.  $u$  is given by

$$u = \frac{C_D^{1-\mu} - 1}{1 - \mu}, \quad \mu > 0 \quad (9)$$

where  $\mu$  is the elasticity of the marginal utility of consumption, and its inverse  $1/\mu$  is the intertemporal elasticity of substitution.  $C_D$  represents consumption in terms of the index  $D$ , and  $C_D = D$  in equilibrium. If the consumer maximizes equations (8) and (9) subject to an intertemporal budget constraint, the spending will be allocated according to:

$$\frac{\dot{C}_D}{C_D} = \frac{1}{\mu} \left[ r - \rho - \frac{\dot{p}_D}{p_D} \right], \quad (10)$$

where  $p_D$  is the price index corresponding to  $D$ :

$$p_D = \left[ \int_0^n p(j)^{-\alpha/(1-\alpha)} dj \right]^{-(1-\alpha)/\alpha}. \quad (11)$$

As mentioned above, the productivity in the R&D activities is assumed to increase with the stock of knowledge  $K_n$ . The cumulative R&D effort increases the stock of knowledge that the researchers have at their disposal. The number of varieties that have been developed can be used as a measure of the cumulative R&D effort, i.e.

$$K_n = n. \quad (12)$$

This implies that the labor market clearing condition (7) can be written as

$$ag + X = L, \quad (13)$$

where  $g = \dot{n}/n$  is the innovation rate of the economy.

The no-arbitrage condition can now be expressed as:

$$\frac{(1-\alpha)X}{\alpha a} = \rho + \beta_D g, \quad (14)$$

where  $\beta_D = 1 + \frac{1+\alpha}{\alpha}(\mu - 1)$ . The left-hand side of equation (14) is the inverse of the price earning ratio, and the right-hand side is the effective cost of capital. These two equations characterize the steady state of this economy.

The steady state is illustrated in Figure 1. The NN curve is the no-arbitrage condition (14), while the LL curve represents the resource constraint (13). The NN curve slopes upwards and illustrates the fact that innovation is driven by the pursuit of profits. An increase in  $g$  leads to a higher cost of capital because the real

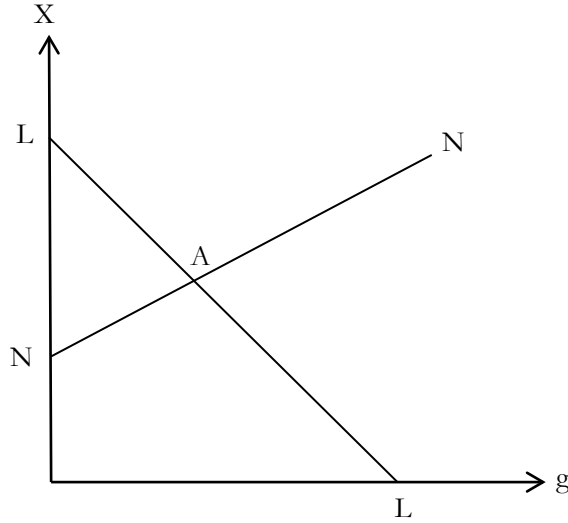


Figure 1: Steady state

interest rate increases and the value of the firm depreciates faster. This implies that the entrepreneur requires a higher profit rate to invest in R&D. The LL curve slopes downwards and reflects that an increase in the rate of innovation requires more employment in R&D, and hence less employment in manufacturing. Long-run equilibrium is characterized by the intersection of these two curves in the point A.

A central feature of the model is that the stock of knowledge does not only affect the productivity in the R&D activities, it also affects the productivity in manufacturing. Grossman and Helpman (1991) use a simple approach to illustrate this. It turns out that the Dixit-Stiglitz index (1) can be interpreted as a production function exhibiting constant returns to scale. In this case, the household consumes a single good, in quantity  $D$ . This final good is produced by competitive firms using differentiated intermediate good or service  $j$  in quantity  $x(j)$ . Investment in R&D increases the number of available varieties of the intermediate inputs, and one unit of labor produces one unit of the intermediate input.

Assume that all intermediate inputs are produced with the same constant-returns-to-scale production function. In this case, all inputs have the same price and therefore the producers use equal quantities of each. This implies that  $D = n^{1/\alpha}x$ , and  $X = nx$  is the total amount of inputs used in the production of the final good. An expression for TFP can be found by calculating final output per unit of input, which



in this case equals

$$\frac{D}{X} = \frac{n^{1/\alpha}x}{nx} = n^{\frac{1-\alpha}{\alpha}}. \quad (15)$$

With  $0 < \alpha < 1$ , this is clearly increasing in  $n$ . This means that TFP in this economy is an increasing function of the number of available inputs. Grossman and Helpman interpret this as gains from specialization, so that when there are more varieties of intermediate inputs, the production of the final good consists of a larger number of finer production processes (Grossman and Helpman, 1991).

So far I've assumed that the intermediate inputs used in the production of the final good are horizontally differentiated. This means that a new input is equally good as an old, and investment in R&D simply increases the number of available varieties. The fact is that inputs can also be vertically differentiated, so that the development of a new input will replace an old one because of its increased quality. Another model also presented in Grossman and Helpman (1991) contains this feature, the so-called *quality ladder* model. They show that the same arguments hold in this case, even though the mechanisms are different. With vertically differentiated inputs, investments in R&D increase the quality of the inputs, which are more productive in manufacturing the final good. Hence, also in this model, R&D investments lead to an increase in TFP (Grossman and Helpman, 1991).

The spillover effects also arise from different sources in the two models. In the case of horizontally differentiated inputs, the spillover effects come from the fact that investment in R&D increases the number of available inputs, which increases the stock of knowledge and decreases future R&D costs. With vertically differentiated inputs, the development of a higher quality input will enable future entrepreneurs to build on a higher quality foundation. This implies that there are spillovers from current to future R&D activities in both cases (Coe et al., 2009).

### 2.2.2 Coe and Helpman's Specification

The model outlined above constitutes the basis for the empirical relationships specified in Coe and Helpman (1995). The variables they use are aggregated to the country level, but the mechanisms are the same. Their specification holds both for horizontally and vertically differentiated inputs. With horizontally differentiated inputs, the number of available inputs is an increasing function of the cumulative R&D expenditure of the country. With vertically differentiated inputs, it is the quality

of the inputs that depend on the cumulative R&D expenditure of the country. In either case, investments in R&D will result in an increase in its TFP.

The model can be extended to include both capital accumulation and international trade. Grossman and Helpman (1991) show that the introduction of capital accumulation does not alter the relationship between R&D investments and TFP. Introducing international trade complicates the model significantly, but Coe and Helpman make the following simple argument: suppose that all intermediate inputs were traded internationally. Then a country's TFP would not depend on its own R&D stock, but on the world's cumulative R&D, because the country would have access to inputs developed as a result of R&D effort in all other countries. This is probably not a realistic scenario, and Coe and Helpman argue that this extreme case is no more likely than the opposite – that no intermediate inputs are traded. They suggest that the most appropriate solution is to allow for something in between, namely that both domestic and international R&D can affect a country's TFP.

The empirical equation relates a country's TFP to domestic and imported R&D in the following way:<sup>8</sup> The output of country  $i$  can be described by the following Cobb-Douglas production function:

$$Y_i = A_i K_i^\alpha N_i^\alpha L_i^{1-\alpha}, \quad (16)$$

where  $A_i$  is a country-specific constant,  $K_i$  is capital services,  $L_i$  is labor services and  $N_i$  is the range of intermediate inputs used in country  $i$ . Taking the logarithms of this expression and rearranging yields:

$$\ln A_i + \alpha \ln N_i = \ln Y_i - \alpha \ln K_i - (1 - \alpha) \ln L_i. \quad (17)$$

This equation exhibits constant returns to capital and labor. TFP defined as:

$$\ln TFP_i = \ln Y_i - \alpha \ln K_i - (1 - \alpha) \ln L_i, \quad (18)$$

will therefore be positively related to the range of intermediate inputs employed:

$$\ln TFP_i = \ln A_i + \alpha \ln N_i. \quad (19)$$

Because Coe and Helpman want to estimate the effect of domestic and foreign R&D separately, they divide  $N_i$  into domestic and foreign inputs,  $N_i^d$  and  $N_i^f$ . Following

---

<sup>8</sup>The following exposition is due to Keller (1998).

the argument presented above, with a certain choice of units,  $N_i^d$  equals the cumulative stock of domestic R&D expenditures, denoted  $S_i^d$ . This domestic R&D stock is constructed for each country using the perpetual-inventory model:

$$S_t^d = (1 - \delta)S_{t-1} + R_{t-1}, \quad (20)$$

where  $R_{t-1}$  is R&D expenditure in year  $t - 1$ , and  $\delta$  is the depreciation rate of the R&D stock. Coe and Helpman use  $\delta = 0.05$ , and argue that their results hold for different values of  $\delta$  (Coe and Helpman, 1995). The R&D stock for the base year is defined as:  $S_0 = R_0/(\delta + g)$ , where  $R_0$  is the R&D expenditure of the base year, and  $g$  is the average growth rate of the R&D stock. In Coe and Helpman this is defined as  $g = \ln(R_{1985}/R_{1970})/15$ .

The foreign input variable  $N_i^f$  is replaced by a measure of a country's imported R&D stock, which is constructed using the weighted sum of the domestic R&D stock of the country's trade partners:

$$S_i^f = \sum_{i \neq j} w_{ij} S_j^d, \quad (21)$$

where  $S_j^d$  is the domestic R&D stock of country  $j$ , and the weight  $w_{ij}$  is the value of the import from country  $j$  as share of country  $i$ 's total imports:  $w_{ij} = M_{ij} / \sum_{i \neq j} M_{ij}$ , and  $M_{ij}$  is the imports from country  $j$  to country  $i$ . With this specification, the import shares equate to 1:  $\sum_{i \neq j} w_{ij} = 1$ .

Coe and Helpman estimate the following equation for the 22 countries they have data on:

$$\ln TFP_i = \alpha_i^0 + \alpha^d \ln S_i^d + \alpha^f \ln S_i^f + \varepsilon_i, \quad (22)$$

where  $\alpha_i^0$  represents country fixed effects to capture country specific effects on productivity that are not captured by the variables in the equation (Coe and Helpman, 1995).

The data used to calculate TFP, such as business sector capital and employment, is mostly taken from the OECD's *Analytical Data Base*, while the estimates of R&D capital stocks are computed using data on R&D expenditures collected from the OECD's *Main Science and Technology Indicators*. The bilateral import shares were calculated using data from the International Monetary Fund's (IMF) *Direction of Trade*. The countries include 21 OECD countries, and Israel, and the sample period

is 1970 to 1990. The data set consists of 440 observations.<sup>9</sup>

---

<sup>9</sup>A more detailed description of their data set is available in the appendix in Coe and Helpman (1995).

## 3 The Data

### 3.1 Description of the Data Sets

The data set used in this thesis consists of accounting data, manufacturing data, trade data and R&D data for Norwegian firms, as well as data on R&D expenditures on sector level for 22 OECD countries.

The manufacturing statistics contain data on all joint-stock companies in the manufacturing industry. This data is collected at the establishment level and then aggregated to the firm level. The firm is in this context defined as “the smallest combination of legal units that is an organizational unit producing goods or services which benefits from a certain degree of autonomy in decision-making, especially for the allocation of its current resources” (Raknerud et al., 2004).<sup>10</sup> This data is then merged with data from the accounting statistics, which is based on information from the balance sheet and income statements for all Norwegian non-financial joint-stock companies. Both data sets are collected by Statistics Norway, and the merged data set is called the *capital database*.<sup>11</sup> In 2001, the firms included in the capital database constituted 80% of the value added and man-hours worked in the manufacturing industry.

The variables in the capital database include for instance investments, operating income, persons employed, intermediate inputs, and sectorial classification. This last variable is based on the NACE revision 1.1.<sup>12</sup> This NACE variable is a five-digit code, which provides a precise classification of the firms. For my purpose, a less narrow definition is suitable, so my division is based on the first two digits of this code. 23 of the two-digit NACE codes are in the manufacturing industry.

This capital database is merged with R&D data for Norwegian firms. This data set is compiled by the Ragnar Frisch Centre for Economic Research, based on information from Statistics Norway.<sup>13</sup> It includes variables like R&D expenditures and persons

---

<sup>10</sup>Raknerud et al. (2004) are quoting the Business Register Recommendations manual, European Communities, 2003, p. 51.

<sup>11</sup>For a detailed description of this data set, see Raknerud et al. (2004).

<sup>12</sup>An overview of the sectors in manufacturing is provided in Appendix C. For more information about the industry codes, see Statistics Norway (2009).

<sup>13</sup>For a detailed description of this data set, see Lien (2007) and Statistics Norway (2004).

employed in R&D activities. R&D is defined as “all creative work undertaken on a systematic basis in order to increase the stock of knowledge, and the use of this stock of knowledge to devise new applications” (Statistics Norway, 2004).

To be able to construct the foreign R&D variable, I have collected data on R&D expenditures on the sector level from the OECD’s *Structural Analysis* (STAN) database.<sup>14</sup> It contains internationally comparable data on different R&D variables for the OECD member countries, as well as a selection of non-member countries. Data on R&D expenditures is available for roughly 22 of the 29 OECD countries (excluding Norway) in the period from 1995 to 2005. Which countries are represented each year varies, depending on the reporting systems of each country.

This data is taken from the *Analytical Business Enterprise Research and Development* (ANBERD) database, in which the data is adjusted by the OECD to correct for inconsistencies in the official data. In addition, for some of the years without data, and for industries where the data is confidential, the OECD have estimated the missing observations. The industrial classification used in the STAN database is the *International Standard Industrial Classification* (ISIC), revision 3.1. This is identical to the NACE codes used in the capital database on the two-digit level.

The R&D expenditures in the OECD data are available in national currency or dollars, in current or constant prices. To get data that is comparable across countries and years, I had the expenditures reported in constant 2000 dollars, transformed using the purchasing power parity (PPP) exchange rates. These exchange rates adjust for differences in cost levels between the countries, and although they do not measure the specific costs related to R&D activities, they provide a more accurate converter than the market exchange rates. PPP exchange rates are used by for instance the OECD when comparing R&D expenditures across countries. The numbers from the OECD data were converted to the Norwegian currency using the OECD PPP exchange rate statistics from 2000. The data used to compute the TFP and R&D stocks for the Norwegian firms is also deflated, so the variables are comparable.

This data is then matched with trade data, which contains data on the imports and exports of Norwegian firms. One disadvantage with the trade data is that the

---

<sup>14</sup>For a more detailed description of this database, see OECD (2010).

industrial classification used here is the *Standard International Trade Classification* (SITC), revision 2. These codes cannot easily be matched with the NACE or ISIC codes, which means that I cannot simply match the imports and exports with the OECD data on R&D expenditures. Since I am not able to identify what sector a firm is importing from, I assume that the firm imports from the same sector as it operates in. This can seem like a strong assumption, but the input-output tables for imported intermediate inputs in the manufacturing sectors in Norway show that it is not.<sup>15</sup> Since the NACE and ISIC codes are identical at the two-digit level, the data on foreign sectorial R&D expenditure can be matched with the firm data from the capital database.

## 3.2 Sample Selection

In the capital database the observations span from 1993 to 2006, while in the R&D data, the observations start in 1993 and end in 2005. The trade data covers 1996 through 2006. I have therefore chosen the period 1996-2005 to include the years where there are observations in all data sets. I also use the R&D observations from 1995 to construct the R&D stock variable for the base year. One issue is that from 1995 to 2001, the R&D data are only sampled every other year. This means that there are no observations for 1996, 1998 and 2000. This will create some trouble when constructing these stock variables, which is explained in further detail in Section 3.3.

All firms with at least 50 employees are included in the R&D data. There are no firms with less than 10 employees in the R&D statistics collected by Statistics Norway. For the firms with 10 to 49 employees, a sample of the firms are drawn each year and the data is estimated based on this sample. This makes the data material for the small firms less reliable and more prone to selection bias. For instance, if the firms that are included in the sample perform more R&D than those that are not included, the sample selection will not be representative. This can cause the estimates to be biased. I have therefore chosen to disregard firms with less than 50 employees.

---

<sup>15</sup>In 2001, for 14 of the 23 sectors, the majority of the imports come from the same sector. For 7 of the 9 remaining sectors, imports from own sector is among the top 3 sources of intermediate inputs. See Appendix D for tables.

Table 1: Descriptive Statistics (2001)

Sector	No. of firms	% with R&D > 0	Mean R&D	Median R&D
15	150	32.0	5.64	2.13
17	23	43.5	1.60	1.10
18	4	75.0	1.73	0.94
20	49	36.7	1.18	1.07
21	27	48.1	11.50	2.46
22	97	11.3	2.32	2.18
24	35	74.3	44.00	19.30
25	25	48.0	2.63	1.47
26	35	60.0	3.25	1.90
27	37	67.6	19.00	2.85
28	77	39.0	2.38	2.05
29	80	61.3	16.30	3.10
31	19	68.4	18.90	9.10
33	19	68.4	35.00	24.50
34	16	81.3	36.60	7.50
35	95	31.6	11.30	2.20
36	49	49.0	3.72	1.97
37	2	50.0	0.38	0.38

Mean and median R&D expenditures apply to R&D performing firms.

Mean and median R&D expenditures are reported in million NOK.

The final data set contains 8188 observations on 1579 Norwegian firms, from 1995 to 2005. There are 3605 observations with positive R&D expenditures, due to 897 firms. Following the procedure in Coe and Helpman (1995) described in Section 3.3, the firm has to be present in the sample in 1995 to have a non-missing R&D stock. There are 1030 observations with positive R&D stocks, due to 124 firms. The data set is not balanced, as there are many firms that are present in only some of the years. 18 of the 23 sectors in manufacturing have firms with positive R&D expenditures, and some descriptive statistics of these sectors are presented in Table 1.

From Table 1, it is evident that there is great variation between the sectors with



regards to the percentage of the firms with R&D investments. 81.3% of the firms in NACE 33, which is “Manufacture of medical, precision and optical instruments, watches and clocks”, have positive R&D expenditures, while that applies to only 11.3% of the firms of NACE 22, “Publishing, printing and reproduction of recorded media”. There are also differences between the sectors with regard to the means of the R&D expenditures of the R&D performing firms. The mean expenditure is highest in NACE 24, “Manufacture of chemicals and chemical products”, and lowest in NACE 37, “Recycling”.

### 3.3 The Variables

#### 3.3.1 TFP

The TFP variable is constructed as follows: Assume a standard Cobb-Douglas production function for firm  $i$ :

$$Y_i = E_i K_i^{\beta_1} L_i^{\beta_2}.$$

Taking the logarithms of this expression yields:

$$y_i = \beta_1 k_i + \beta_2 l_i + e_i,$$

where  $y_i = \ln Y_i$ , etc. This expression is used to get estimates of TFP, given as:

$$e_i = y_i - \hat{\beta}_1 k_i - \hat{\beta}_2 l_i.$$

Estimating this with ordinary least squares (OLS) would in most cases give biased results because of the endogeneity of the factor inputs and TFP. Say that TFP for some reason rises, for instance because of successful process innovation, and the firm can observe this prior to setting the amount of labor and capital to use in the production. Then the firm can decide to adjust the amount of labor and/or capital, and this will make the factor inputs correlated with TFP. If any variable used as a regressor is correlated with the error term, the OLS estimates will be biased.<sup>16</sup> To avoid this problem, TFP is estimated using the Olley-Pakes (1996) method, which gives estimates that are robust to endogeneity.<sup>17</sup>

---

<sup>16</sup>This is explained in Section 4. See Kennedy (2008) for further details.

<sup>17</sup>It is beyond the scope of this thesis to explain this method in detail. See Olley and Pakes (1996), or Ekholm et al. (2009) for the exact procedure used in my data set.

### 3.3.2 R&D stock

The R&D stock variable for the firms is constructed in the same way as in Coe and Helpman (1995), using the perpetual inventory method:

$$S_t^d = (1 - \delta)S_{t-1} + R_{t-1},$$

where  $R_{t-1}$  is R&D expenditure in year  $t - 1$ , and  $\delta$  is the depreciation rate of the R&D stock. I follow Coe and Helpman and use  $\delta = 0.05$ , and the stock variable for the base year (1995) is given by  $S_{1995} = R_{1995}/(\delta + g)$ . The growth rate  $g$  is defined by  $g = \ln(R_{2005}/R_{1995})/10$ .

For some firms, the R&D stock variable for the base year is negative. This occurs when the firm invested so much more in R&D in 1995 than in 2005 that the growth rate is negative by more than 0.05. Then the denominator of the equation for the base year is negative, which makes the stock for the base year negative. Having a negative R&D stock is counterintuitive. In those cases, I reset the growth rate to zero and constructed the stock on basis of the R&D expenditure in 1995 and the depreciation rate. Since this gives a somewhat higher value for the stock than would be the case if the firm had a positive growth rate, this approach will overestimate the R&D stocks of these firms compared to the other firms. I also tried resetting the R&D stock variable to zero in the base year, which would underestimate the stock compared to the other firms. Both of these approaches lead to negligible changes in the results.

As mentioned in Section 3.1, the Norwegian R&D data are only sampled every other year from 1995 to 2001. To be able to generate the R&D stock variables, the missing observations for 1996, 1998 and 2000 have to be estimated. Coe and Helpman estimate their missing R&D expenditure observations by regressing the firm's R&D expenditure on its investment and value added, all in logarithms, and then use this to predict the missing observations. I use the same approach, and I also tried calculated the average of the R&D expenditure of the previous and the following year, and using this as an estimate for the missing observations. Both procedures give roughly the same estimates.

### 3.3.3 Foreign R&D stock

The firm's imported R&D stock  $S_i^f$  is constructed using the weighted sum of the R&D stock of the corresponding sector in the countries that the firm imports from:

$$S_i^f = \sum_{i \neq j} w_{ij} S_j^d,$$

where  $S_j^d$  is the R&D stock of sector  $s$  in country  $j$ , constructed in the same way as the firm's own R&D stock.<sup>18</sup> The weight  $w_{ij}$  is the value of the import from country  $j$  as share of firm  $i$ 's total imports:  $w_{ij} = M_{ij} / \sum_{i \neq j} M_{ij}$ .  $M_{ij}$  is the imports from country  $j$  to firm  $i$ .

---

<sup>18</sup>The sector subscript is dropped for simplicity.

## 4 Empirical Strategy and Results

I reproduce the results from Coe and Helpman (1995) using the data set described above. This is *panel data*, which means that each individual (firm) is observed in different time periods. This kind of data has two sources of variation, *within* firm and *between* firm variation. Figure 2 illustrates the difference between these two types of variation.<sup>19</sup> The observations in each ellipses belong to a single firm, and this is the within firm variation. The variation from ellipse to ellipse is the between firm variation.

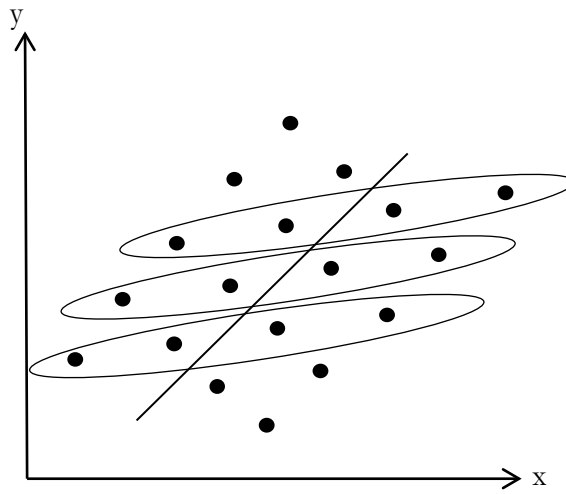


Figure 2: Within and between variation

Panel data requires some special estimation techniques, as using simple OLS would produce biased estimates in most cases. If there exists a variable that influences the dependent variable, but is left out of the equation, this will be incorporated in the error term. If this omitted variable is correlated with the included explanatory variables, there will be correlation between the error term and the regressors.

Then if an included explanatory variable which is positively correlated with the error term increases, the dependent variable will increase for two reasons. It will increase because of the direct effect through the independent variable, and because of the indirect effect through the error term. The OLS estimator will attribute both these effects to the independent variable, and give a biased estimate of the coefficient.

---

<sup>19</sup>The exposition is due to Kennedy (2008).

This is called an *omitted variable bias*. There are so many different factors affecting a firm's productivity, it is unlikely that the specification estimated by Coe and Helpman contains all the relevant variables. This suggests that the OLS estimator is likely to give biased estimates in this case.

One way of avoiding this problem is by using the so-called *fixed effects* estimator. This is essentially the same as using OLS on the deviations from the mean for each firm, which is beneficial because it controls for any time-invariant, unobserved heterogeneity that affects the firm's productivity. This estimator only uses the variation within each firm, and is also called the *within estimator*.

Suppose the equation to estimate is of the form:

$$y_{it} = \beta_1 x_{it} + \varepsilon_{it},$$

where the error term  $\varepsilon_{it}$  is composed by a firm-specific error  $v_i$  and a regular error  $u_{it}$ . The equation can be rewritten as:

$$y_{it} = \beta_1 x_{it} + v_i + u_{it}.$$

Here, there is some time-constant, unobservable variable that affects the dependent variable for each firm  $i$ . By subtracting the average for each firm, this time-invariant variable is removed:

$$y_{it} - \bar{y}_i = \beta_1 (x_{it} - \bar{x}_i) + u_{it} - \bar{u}_i.$$

The fixed effects estimator is identical to running OLS on this expression, except that the sum of squared errors has to be divided by  $NT - K - N$  instead of  $NT - K$ , because there are  $N$  estimated means.<sup>20</sup> This procedure is equivalent to including a dummy variable for each firm, which allows each firm to have its own intercept.

There is another way of allowing for different intercepts, which is called the *random effects* model. This estimator is more efficient than both OLS and fixed effects, because it uses both the within and the between variation. It does, however, have one major drawback. If the intercepts are correlated with the explanatory variable, the random effects estimator is biased. This is because it interprets the different intercepts as randomly drawn, and treats them as part of the error term. This implies that the error term will be correlated with the explanatory variable, which creates

---

<sup>20</sup>See Kennedy (2008).

biased estimates, as explained above.

Figure 2 illustrates panel data with positive correlation between the intercept and the explanatory variable. Estimating the relationship between  $x$  and  $y$  with OLS would produce the slope estimate illustrated by the straight line. The OLS estimator is overestimating the slope, because it does not recognize that the observations inside the ellipses belong to the same firm. The random effects estimator will also create a biased estimate in this case, for the reason explained above. The fixed effects estimator, on the other hand, explicitly includes dummy variables for the different intercepts. Since the specification in Coe and Helpman (1995) most likely does not include all the variables that affect a firm's productivity, as mentioned above, the fixed effects estimator is the most suited for this type of estimation. Coe and Helpman let the constant term vary for each country and use the OLS estimator, which is equivalent to using the fixed effects estimator.

I estimate the most basic specification in Coe and Helpman (1995):

$$\ln TFP_{it} = \alpha_i + \alpha^d \ln S_{it}^d + \alpha^f \ln S_{it}^f + \varepsilon_{it}, \quad (23)$$

where  $S_i^d$  is firm  $i$ 's R&D stock in time  $t$ , and  $S_i^f$  is the weighted sum of the R&D stock of the corresponding sector in the countries that firm  $i$  imports from.  $\alpha_i$  represents firm fixed effects. For reference, I also estimate firm TFP as a function of its own R&D stock only:

$$\ln TFP_{it} = \alpha_i + \alpha^d \ln S_{it}^d + \varepsilon_{it}. \quad (24)$$

The results of these regressions are presented in Table 2. These results indicate that the firm's own R&D stock has a positive and significant effect on the productivity of the firm, as predicted by the theory. When including the foreign R&D stock, this positive effect of the firm's own R&D is maintained. The foreign R&D stock variable, however, is not significantly different from zero at any of the usual significance levels. This indicates that at the firm level, imported R&D does not affect productivity. The size of the estimated coefficient on the firm-level R&D stock is smaller than the estimated coefficient for the domestic R&D stock in Coe and Helpman (1995), which is 0.097. The estimated coefficient for the foreign R&D stock is 0.0075, which is markedly smaller than 0.092, which is the estimate in Coe and Helpman (1995).

Table 2: Regression 1a

	(1)	(2)
R&D stock	0.0572** (0.0236)	0.0640*** (0.0239)
Foreign R&D stock		0.0075 (0.0109)
<i>N</i>	1080	1020

\*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Standard errors in parentheses.

The effects of the domestic and the foreign R&D stocks are almost equal in magnitude in Coe and Helpman (1995). This indicates that on the country level, the country benefits almost as much from spillovers from their trade partners' R&D effort as from their own. At least this applies to the countries in the sample used in the paper, which all invest in domestic R&D. As mentioned before, this is seen as a prerequisite for a country to be able to benefit from imported R&D.

On the firm level, however, the imported foreign R&D stock appears to have no significant impact at all on the firm's productivity. This indicates that any spillover effects are not transmitted through each firm's trade in intermediate inputs. It might be that spillover effects would be detected at a more aggregate level, for instance when comparing industries. That the coefficient of the firm-level R&D stock is smaller than the coefficient of the country-level R&D stock, could also indicate that there are domestic spillover effects that only affect the productivity at the aggregate level.

#### 4.1 Clustered Standard Errors

In the regression results shown in Table 2, the standard errors are unadjusted. This can be problematic, because the error term might be heteroskedastic, as well as subject to intra-firm correlation. If the error term suffers from heteroskedasticity, this means that the variance of the error term varies with the independent variable, for instance if the variance of the error term increases as the R&D stock increases. If there is intra-firm correlation, the error term for firm  $i$  in year  $t$  will be correlated

with the error term for firm  $i$  in the other periods. Suppose that some shock hits the firm in one year, which affects the productivity of the firm in the following years. If the effects of the shock is reduced over time, it will not be incorporated in the time-invariant variable, but in the error term. This will cause the error terms to be serially correlated. In either case, this will not affect the estimated coefficients, but it will bias the standard errors used for inference.

To correct for this, the standard errors need to be *clustered* on firms. Clustering produces standard errors that are robust to both heteroskedasticity and intra-firm correlation.<sup>21</sup> The clustering procedure can be illustrated as follows:<sup>22</sup>

Suppose, for simplicity, the model is of the form:

$$y_{ig} = \beta_0 + \beta_1 x_i + e_{ig},$$

where  $y_{ig}$  is the dependent variable for individual  $i$  in group  $g$ , and the explanatory variable  $x_i$  varies only at the group level. Assume that the error terms are correlated within groups:

$$E[e_{ig}e_{jg}] = \rho_e \sigma_e^2 > 0,$$

where  $\rho_e$  is the so-called *intraclass correlation coefficient*. Assume further that the error term has the following structure:

$$e_{ig} = \nu_g + \eta_{ig},$$

where  $\nu_g$  is the group term and  $\eta_{ig}$  is a regular error term, which is uncorrelated. The intraclass correlation coefficient can then be written:

$$\rho_e = \frac{\sigma_\nu^2}{\sigma_\nu^2 + \sigma_\eta^2},$$

where  $\sigma_\nu^2$  is the variance of  $\nu_g$  and  $\sigma_\eta^2$  is the variance of  $\eta_{ig}$ . An expression can be found that shows how much the estimated variance is biased:

$$\frac{V(\hat{\beta})}{V_c(\hat{\beta})} = 1 + (n - 1)\rho_e,$$

where  $V(\hat{\beta})$  is the ordinary OLS variance formula, the correct variance that applies to the given error structure is given by  $V_c(\hat{\beta})$ , and  $n$  is the size of the groups. This

---

<sup>21</sup>See Angrist and Pischke (2009) and Wooldridge (2003).

<sup>22</sup>This illustrative example is borrowed from Angrist and Pischke (2009).



expression shows that if the intraclass correlation coefficient or the group size is high, the default standard errors will be heavily biased. The square root of this expression is called the *Moulton factor*.

The assumption in the example above that the explanatory variable varies only at the group level, is included to simplify the exposition. The Moulton factor can be extended to allow the explanatory variable to vary at the individual level, i.e. vary in each time period. The more general formula also applies to data with different group sizes, i.e. with different number of observations per firm. The general Moulton factor is the square root of this expression:

$$\frac{V(\hat{\beta})}{V_c(\hat{\beta})} = 1 + \left[ \frac{V(n_g)}{\bar{n}} + \bar{n} - 1 \right] \rho_x \rho_e,$$

where  $V(n_g)$  is the variance structure of the group size  $n$ ,  $\bar{n}$  is the average size of the groups, and  $\rho_x$  is the intraclass correlation of the regressor, given by:

$$\rho_x = \frac{\sum_g \sum_j \sum_{i \neq j} (x_{ig} - \bar{x})(x_{jg} - \bar{x})}{V(x_{ig}) \sum_g n_g(n_g - 1)},$$

where  $V(x_{ig})$  is the variance structure of the regressor. The expression for the general Moulton factor shows that the bias of the default standard errors is larger when the group size varies, and when the intraclass correlation of the regressor or the residual is high.

The cluster covariance matrix is given by:

$$\hat{\Omega}_{cl} = \mathbf{X}'\mathbf{X}^{-1} \left( \sum_g \mathbf{X}_g \hat{\Psi}_g \mathbf{X}_g \right) \mathbf{X}'\mathbf{X}^{-1},$$

where  $\hat{\Psi}_g$  is the covariance matrix of the within-group residuals, including a degrees of freedom adjustment. The clustered estimator is consistent given any within-group correlation structure, and clustering provides standard errors that are robust to both heteroskedasticity and intra-firm correlation.

The clustered standard errors are likely to be higher than the non-adjusted standard errors. The reason is that instead of treating each observation as independent within each firm, the estimator now controls for the fact that the R&D stock for firm  $i$  in

year  $t$  is likely to be correlated with the R&D stock for firm  $i$  in year  $t + 1$ . It also allows the error term within firms to be correlated over time. This reduces the amount of independent variation in the data, which increases the standard errors.

Table 3: Regression 1b

	(1)	(2)
R&D stock	0.0572*	0.0640**
	(0.0299)	(0.0294)
Foreign R&D stock		0.0075
		(0.0141)
$N$	1080	1020

\*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Standard errors in parentheses.

With panel data data, the group  $g$  is the cross-sectional unit (firm), and the individual  $i$  is the time (year). In Table 3, the standard errors are clustered on firms. The standard errors are slightly higher for both variables than without clustering, as expected. The R&D stock variable is still significant at the 10% level in the specification without the imported R&D stock variable. When including this variable, the coefficient becomes significant at the 5% level, as it was with the default standard errors. The foreign R&D stock variable is still not significant at any of the conventional levels.

The clustering procedure is sensitive to the number of clusters in the data set. If there are too few clusters, the serial or intraclass correlation could be underestimated. Angrist and Pischke (2009) suggest that the estimates are trustworthy when the data contains at least 50 clusters. In my data set there are 124 clusters, so the inference should be reliable.

## 4.2 Extension I: Nonstationarity

On the aggregated level used by Coe and Helpman, TFP, R&D stock and foreign R&D stock are all nonstationary. This is common for macroeconomic data, and implies that OLS can give so-called *spurious* results. The test statistics such as the  $R^2$ , DW and  $t$  statistics, are not valid under nonstationarity. If these are used for

inference, relationships between variables may incorrectly appear to be statistically significant.

In Coe and Helpman (1995), they use OLS and let the constant term differ for each country, which is essentially the same as using the fixed effects estimator. They discuss the fact that the variables are nonstationary and they rely on the method of *cointegration*. Consider a first order autoregressive model of the form:  $y_t = \alpha y_{t-1} + \varepsilon_t$ , where  $\varepsilon_t$  is a stationary error term. If  $\alpha = 1$ , the variable  $y$  is determined by last period's value, plus a random error, and the variable is said to have a *unit root*. By differencing this variable, you get  $y_t - y_{t-1} = \varepsilon_t$ , which is stationary. A nonstationary variable that can be made stationary by taking first differences is said to be integrated of order 1, written  $I(1)$ .

Suppose that there are two nonstationary variables that are  $I(1)$ , but that the relationship between those two variables is stationary. The idea is that even though the two variables both trend upwards, the distance between them does not. The variables are cointegrated, and regressing one on the other will give stationary residuals. Textbook examples include prices and wages, household income and expenditure, and imports and exports. If a set of variables are cointegrated, the regression estimates will not be spurious. In addition, while estimates from cointegration regression are biased in small samples, they are *superconsistent* in large samples. Superconsistency implies that the estimates converge on the true parameter value faster than with stationary variables.<sup>23</sup>

In Coe and Helpman (1995), the estimates are interpreted as pooled cointegrated equations. No standard errors are presented because “they are, in general, biased and their distribution is not asymptotically normal” (Coe and Helpman, 1995). In a follow-up paper from 2009, Coe, Helpman and Hoffmaister perform the same regressions with an extended data set and new econometric methods. They perform tests on the variables to confirm that they are nonstationary and cointegrated, and then use dynamic OLS to estimate long-run relationships. Dynamic OLS is used to get both pooled (within) and group mean (between) estimates. The estimates reported in this paper confirm the original results.<sup>24</sup>

---

<sup>23</sup>For more on cointegration, see Kennedy (2008) and Coe et al. (2009).

<sup>24</sup>See Coe et al. (2009).

However, cointegration techniques cannot be utilized unless the variables are actually nonstationary and cointegrated, which typically applies to macroeconomic data. There are several so-called *panel unit root tests* that can be applied to test whether or not a variable has a unit root, and hence is nonstationary. For this type of panel data, with few time periods and a relatively large number of panels, the Harris-Tzavalis test is best suited. This test has as the null hypothesis that all the panels have a unit root, and it allows for fixed effects.<sup>25</sup> The results of this test is presented in Table 4, and shows that the TFP and foreign R&D stock variables do not have a unit root. The firm-level R&D stock is however still nonstationary, by construction. This could indicate that the positive effect of the firm's own R&D stock is a spurious result.

Table 4: Harris-Tzavalis test

	Test statistic
TFP	0.3460***
R&D expenditure	0.4623***
R&D stock	0.9338
Foreign R&D stock	0.3109***

\*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

The null hypothesis of a unit root is rejected if the test statistic is significant.

So with firm-level data, it seems that the TFP and foreign R&D stock variables are no longer nonstationary. This is quite intuitive. While for a country as a whole, the productivity might be steadily rising, this need not apply to each firm. For a single firm it is likely that the productivity can fluctuate in intervals that differ from firm to firm, or sector to sector. When all these firms are aggregated, these fluctuations are hidden by the overall trend, which is positive.

One way of avoiding these complications is to use the differenced variables and look at the growth rates of the R&D stocks in relation to the change in TFP. This is not done by Coe and Helpman, because they want to estimate the long-run relationship. In addition, Klette and Kortum summarize some empirical findings in the litera-

<sup>25</sup>For more details on this test, see Harris and Tzavalis (1999).

ture, and one of their stylized facts is the following: “The longitudinal (within-firm, across-time) relationship between firm-level differences in R&D and productivity growth, which controls for permanent differences across firms, has turned out to be fragile and typically not statistically significant” (Klette and Kortum, 2004). I would therefore not expect to find any significant effect of changes in either of the R&D stocks on the productivity growth of the firms.

Therefore, I have tried another approach, and replaced the R&D stock variable with the firm’s lagged R&D expenditure as a robustness check. The equation to estimate is then given by:

$$\ln TFP_{it} = \alpha_i + \alpha^d \ln R_{it-1}^d + \alpha^f \ln S_{it}^f + \varepsilon_{it}, \quad (25)$$

where  $R_{it-1}^d$  is firm  $i$ ’s R&D expenditures in the previous period. This is contrary to the theory presented in chapter 2.2, which stresses that the relevant variable is the stock of knowledge. However, on the firm-level, there might be arguments why lagged R&D expenditure can work as a substitute for the stock. For a country, the cumulative R&D stock is the relevant variable, because through the cumulative R&D performance, the country gets a wider range of (horizontally differentiated) or higher quality (vertically differentiated) inputs to choose from. For the firm, the R&D effort exerted eight or ten years ago, might not in the same way affect productivity today.

Table 5: Regression 2

	(1)
Lagged R&D expenditure	0.0256** (0.0124)
Foreign R&D stock	-0.0007 (0.0091)
$N$	2606

\*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Standard errors in parentheses.

The results from this regression is presented in Table 5, with standard errors clustered on the firm level. The number of observations has increased notably, because firms with just a single observation on own R&D expenditure now can be included

in the sample, if they have imported from at least one of the countries included in the OECD data. There is still a positive and significant effect of the R&D stock, as expected. However, the size of the estimated coefficient for the lagged R&D expenditures is smaller than it was for the R&D stock. This could indicate that some of the effect previously attributed to the R&D stock is indeed spurious. Nevertheless, using lagged R&D expenditure as an explanatory variable is not without its drawbacks. It is hard to justify why the effect of a R&D investment should occur in the following year, and not two, three or several years later. In this aspect, the stock variable is more suitable. I will therefore continue using the stock variable, and be careful when interpreting the coefficient, as it may be over-estimated.

The foreign R&D stock variable is still not significant at any of the conventional levels, and the coefficient is now negative. This alternative specification has not revealed any effect of imported R&D spillovers.

### 4.3 Extension II: Different weights

As previously mentioned, Coe and Helpman have been criticized for their use of bilateral import shares as weights when constructing the foreign R&D stock variable. The endogeneity of the trade flows, I have already argued, is not a concern with firm-level data. However, a firm can – unlike a country – purchase intermediate inputs domestically as well as internationally. An alternative approach therefore consists of replacing the import share weights with the value of the import from country  $j$  as share of the value of firm  $i$ 's total purchase of intermediate inputs. By reflecting the fact that a firm can trade with both domestic and foreign firms, the importance of the imported inputs is reduced. Also, the relevance of the foreign R&D stock variable is higher for firms with a large share of imported inputs in this specification.

The estimating equation still has the original form:

$$\ln TFP_{it} = \alpha_i + \alpha^d \ln R_{it}^d + \alpha^f \ln S_{it}^f + \varepsilon_{it}, \quad (26)$$

except that the foreign R&D stock variable now is constructed as:

$$S_i^f = \sum_{i \neq j} z_{ij} S_j^d,$$

where the weight  $z_{ij}$  is the value of the import from country  $j$  as share of firm  $i$ 's total inputs:  $z_{ij} = M_{ij} / \sum_i I_i$ , and  $M_{ij}$  is the imports from country  $j$  to firm  $i$  and  $\sum_i I_i$  is the value of country  $i$ 's total purchase of intermediate inputs.

Table 6: Regression 3

	(1)
R&D stock	0.0597** (0.0287)
Foreign R&D stock	0.0127 (0.0091)
$N$	1018

\*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Standard errors in parentheses.

The results of this regression are presented in Table 6, with the standard errors clustered on firm level. The firm's own R&D stock is still positively related to the firm's productivity, and significant at the 5% level. The foreign R&D stock variable appears to have no influence on the productivity, although the magnitude of the estimated coefficient has increased slightly compared to the previous specifications. This increased coefficient might indicate that the foreign R&D affects the firm more the larger the share of imports in total purchase of intermediate inputs, i.e. the more "open" the firm is to international trade.

However, the effect of the foreign R&D stock is still not significant at any of the conventional levels. Thus, it seems as though the choice of weights only affect the results in a negible way.

#### 4.4 Extension III: Aggregating the foreign R&D stock variable

So far, no significant effect of foreign R&D stock has been detected at the firm level. However, it could be that the positive spillover effects that Coe and Helpman (1995) find on the country level do not come from each firms imports, but that the import of intermediate inputs by one firm transfers knowledge that is dispersed to other firms in the same sector, and thereby increases their productivity. One approach to

investigate this is by aggregating the foreign R&D stock variable to the sector level, and relating the firm's productivity to the firm's own R&D stock and to the total imported R&D stock of the sector that the firm belongs to.

The foreign R&D stock now varies only at the sector level. Hence, the standard errors need to be clustered at this level. If not, the estimator would treat the regression as though all the variables contained variation on the firm level. Quoting Angrist and Pischke: "Making a data set larger by copying a smaller one  $n$  times generates no new information" (Angrist and Pischke, 2009).

The equation to estimate becomes:

$$\ln TFP_{ist} = \alpha_i + \alpha^d \ln S_{ist}^d + \alpha^f \ln S_{st}^f + \varepsilon_{ist}, \quad (27)$$

where  $\ln S_{st}^f$  is the domestic R&D stock of all the countries that sector  $s$  imports from, weighted by either the import share or the share of intermediate inputs purchased by sector  $s$ .

Table 7: Regression 4

	(1)	(2)
R&D stock	0.0579** (0.0239)	0.0576** (0.0236)
Foreign R&D stock	0.0091 (0.0246)	0.0112 (0.0211)
$N$	1036	1036

\*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Standard errors in parentheses.

The results of this regression are presented in Table 7. In column (1), the weight is the import share, while in column (2), it is the imports as share of inputs. In both specifications, the standard errors are clustered at the sector level. There is still no significant effect of the foreign R&D stock variable, while the coefficient for the firm's own R&D stock is of similar magnitude as in the previous specifications. It is still significant at the 5% significance level.



As mentioned earlier, the clustering procedure is sensitive to the number of clusters in the data set. With the aggregated foreign R&D stock variable, there are only 18 clusters in the data set, as there are only 18 NACE codes in manufacturing with positive R&D expenditures. This makes the standard error estimates in Table 7 prone to bias. There are methods to correct for this, however, implementing these is beyond the scope of this thesis. The problem that arises with few clusters, as explained in Section 4.1, is that the serial or intraclass correlation could be underestimated. This implies that the standard errors above, if they are biased, are too small.

Since the coefficient of the foreign R&D stock variable would not be significant with even higher standard errors, I conclude that no effect of the foreign R&D stock is detected in this specification either. That implies that if such spillovers from trade exist, they are not intra-firm or intra-industry, but must appear at a more aggregated level.

## 4.5 Remarks

I have estimated several extensions of the simplest specification in Coe and Helpman (1995), without detecting any significant effect of the foreign R&D stock variable. Coe and Helpman extend their estimating equation in two ways. The first is by weighing the foreign R&D stock variable with the imports as share of GDP, to let the level of imports be reflected. In this way, they can investigate whether countries that are more open to trade benefit more from the imported R&D than other countries. The second is by interacting the domestic R&D stock with a dummy variable that equals one for the seven largest economies compared to the 15 other countries. This is done to let the impact of the domestic R&D variable differ for the countries with the highest GDP, because these countries might be more capable of exploiting available complementarities (Coe and Helpman, 1995). This relates to the idea that the absorptive capacity is important for a country to be able to benefit from foreign R&D.

The first of these extensions is analogous to the input weights used in Section 4.3, where the foreign R&D stock variable has a bigger impact on firms with a high share of imported inputs relative to domestically purchased inputs. This measures how “open” the firm is to international trade. I have also tried several other specifica-

tions. In addition to the three extensions explained here, I also estimated a two-way fixed effects version of the basic specification, in which I allowed for sector-specific time-invariant effects. None of these extensions have revealed any significant effect of imported R&D spillovers.

This does obviously not imply that there are no international R&D spillovers from trade. It merely indicates that the transmission channel for such spillovers is not each firm's imports of intermediate inputs. It can very well be that there exists spillovers from trade at a more aggregate level. It might also be that international trade is not the most important channel for R&D spillovers.

The role of international trade in intermediate inputs in transmitting R&D spillovers has been disputed in the literature. While Coe and Helpman (1995) and Coe et al. (1997) find significant R&D spillovers from international trade, Griffith et al. (2004) and Cameron et al. (2005) find that trade has a positive effect on productivity growth through faster technology transfer, but does not increase rates of innovation. As mentioned in Section 2.1, FDI, international R&D collaborations, patent sales, licencing agreements, joint ventures and migration of researchers are among the other transmission channels suggested to be important for the transfer R&D spillovers.

## 5 Conclusion

In this thesis, I have estimated the basic specification from Coe and Helpman (1995), as well as a number of extensions, using a panel data set on Norwegian firms. The basic regression links a firm's productivity to its own R&D stock and a foreign R&D stock, consisting of a weighted sum of the domestic R&D stock of the firm's trade partners. I use the fixed effects estimator to allow for unobserved heterogeneity between the firms.

The estimated coefficient of the firm's own R&D stock is positive and significant, which implies that the firm's productivity is positively related to its R&D stock. The estimated coefficient of the foreign R&D stock is however small and not statistically significant. This indicates that any spillover effects from trade are not transmitted through each firm's import of intermediate inputs. The first adjustment was to cluster the standard errors, to allow for heteroskedasticity and intra-firm correlation. This procedure does not change the estimates, but increases the standard errors. The extensions to the standard specification estimated in this thesis all have clustered standard errors.

The first extension was to try substituting the firm's own R&D stock with the lagged R&D expenditure, as the R&D stock is a non-stationary variable. This lowered the magnitude of the coefficient, compared to the basic specification, but it is still significant at the 5% level. The foreign R&D stock variable appeared to still have no influence on the firm TFP. The second extension consisted of using different weights, namely the imported inputs as share of the firm's total purchase of intermediate inputs. This extension was made to reflect the fact that firms can purchase inputs domestically as well as internationally, but led to negligible changes in the results. The last extension was aggregating the foreign R&D stock variable to the sector level. The results from this regression confirmed the positive and significant influence of the firm's own R&D stock, while no effect was detected from the aggregate imported R&D stock.

While the effect of the firm's own R&D has been positive and significant in all the specifications presented in this thesis, I have not been able to detect any effect of the foreign R&D stock variable. So the positive spillover effect of imported R&D that is found on the country level in Coe and Helpman (1995), is not found on the

firm level with my data set. Seeing as how a country's productivity is made up by aggregating the firms' productivity, one would think that any spillover effects from trade contributing to the country's productivity would also show up at the firm level.

As mentioned in the introduction, Griliches claims that true spillovers cross the industry boundaries: "True spillovers are ideas borrowed by research teams of industry  $i$  from the research results of industry  $j$ . It is not clear that this kind of borrowing is particularly related to input purchase flows. The photographic equipment industry and the scientific instruments industry may not buy much from each other but may be, in a sense, working on similar things and hence benefiting much from each other's research" (Griliches, 1992).

It is not hard to imagine how such transactions could occur within a firm or an industry as well. Reverse engineering could, for instance, provide the firm with knowledge from an imported intermediate input beyond what the input was intended for. Also, the categories in the NACE industrial classifications are wide enough to include research teams operating in dissimilar fields, which would allow for spillovers of the kind Griliches describes. Still, no such effect can be found in the various specifications estimated in this thesis.

## References

- Aghion, P. and P. Howitt (1992). A Model of Growth Through Creative Destruction. *Econometrica* 60(2), 323–351.
- Angrist, J. D. and J.-S. Pischke (2009). *Mostly Harmless Econometrics*. Princeton University Press.
- Bernstein, J. I. and P. Mohnen (1998). International R&D spillovers between U.S. and Japanese R&D intensive sectors. *Journal of International Economics* 44, 315–338.
- Cameron, G., J. Proudman, and S. Redding (2005). Technological Convergence, R&D, Trade and Productivity Growth. *European Economic Review* 49, 775–807.
- Cincera, M. and B. van Pottelsberghe de la Potterie (2001). International R&D spillovers – a survey. *Cahiers Economiques de Bruxelles* 169(1).
- Coe, D. T. and E. Helpman (1995). International R&D spillovers. *European Economic Review* 39(5), 859–887.
- Coe, D. T., E. Helpman, and A. W. Hoffmaister (1997). North-South R&D Spillovers. *The Economic Journal* 107, 134–149.
- Coe, D. T., E. Helpman, and A. W. Hoffmaister (2009). International R&D Spillovers and Institutions. *European Economic Review* 53(7), 723–741.
- Coe, D. T. and A. W. Hoffmaister (1999). Are There International R&D Spillovers Among Randomly Matched Trade Partners? A Response to Keller. *IMF Working Paper* (18).
- Cohen, W. M. and D. A. Levinthal (1989). Innovation and Learning: The Two Faces of R & D. *The Economic Journal* 99(397), 569–596.
- Eaton, J. and S. Kortum (1996). Trade in Ideas – Patenting and Productivity in the OECD. *Journal of International Economics* 40, 251–278.
- Ekholm, K., A. Moxnes, and K. H. Ulltveit-Moe (2009). Manufacturing Restructuring and the Role of Real Exchange Rate Shocks. *CEPR Discussion Papers* (6904).
- Griffith, R., S. Redding, and J. V. Reenen (2004). Mapping the Two Faces of R&D: Productivity Growth in a Panel of OECD Industries. *The Review of Economics and Statistics* 86(4), 883–895.

- Griliches, Z. (1992). The Search for R&D Spillovers. *The Scandinavian Journal of Economics* 94, 29–47.
- Grossman, G. and E. Helpman (1991). *Innovation and Growth in the Global Economy*. MIT Press.
- Harris, R. D. and E. Tzavalis (1999). Inference for unit roots in dynamic panels where the time dimension is fixed. *Journal of Econometrics* 97, 201–226.
- Helpman, E. (1992). Endogenous Macroeconomic Growth Theory. *European Economic Review* 36, 237–267.
- Jaffe, A. B. (1986). Technological Opportunity and Spillovers of R&D – Evidence from Firms’ Patents, Profits and Market Value. *American Economic Review* 75(6), 984–1002.
- Jaffe, A. B., M. Trajtenberg, and R. Henderson (1993). Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *The Quarterly Journal of Economics* 108(3), 577–598.
- Keller, W. (1998). Are International R&D Spillovers Trade-related? Analyzing Spillovers Among Randomly Matched Trade Partners. *European Economic Review* 42(8), 1469 – 1481.
- Kennedy, P. (2008). *A Guide to Econometrics* (6 ed.). Blackwell Publishing Ltd.
- Klette, T. J. and S. Kortum (2004). Innovating Firms and Aggregate Innovation. *Journal of Political Economy* 112(5), 986–1018.
- Lichtenberg, F. R. and B. van Pottelsberghe de la Potterie (1996). International R&D Spillovers – a Re-examination. *NBER Working Paper Series* (5668).
- Lien, G. (2007). Dokumentasjon av tidsseriedatabase for FoU-data. Statistics Norway. Notater 2007/55.
- Lucas, R. E. (1988). On the Mechanics of Economic Development. *Journal of Monetary Economics* 22(1), 3 – 42.
- Nadiri, M. I. (1993). Innovations and Technological Spillovers. *NBER Working Paper Series* (4423).
- OECD (2010). Structural Analysis Database. <http://www.oecd.org/sti/stan/>.

- Olley, G. S. and A. Pakes (1996). The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica* 64(6), 1263–1297.
- Raknerud, A., D. Ronningen, and T. Skjerpen (2004). Documentation of the Capital Database. *Statistics Norway Documents 2004/16*, 1–13.
- Romer, P. M. (1986). Increasing Returns and Long-Run Growth. *The Journal of Political Economy* 94(5), 1002–1037.
- Romer, P. M. (1990). Endogenous Technological Change. *The Journal of Political Economy* 98(5), 71–102.
- Statistics Norway (2004). Research and Development in the Business Enterprise Sector 2001-2002. *Official Statistics of Norway D 301*.
- Statistics Norway (2009). Standard Industrial Classification (SIC2002). <http://www3.ssb.no/stabas/ItemsFrames.asp?ID=5556001&Language=en>.
- Wooldridge, J. M. (2003). Cluster-Sample Methods in Applied Econometrics. *The American Economic Review Papers and Proceedings of the One Hundred Fifteenth Annual Meeting of the American Economic Association, Washington, DC, January 3-5, 2003* 93(2), 133–138.

## Appendix A

Table 8: OECD R&D measures

	Gross R&D expenditure per capita	R&D personnel/ total employment
Norway	590.026	11.627
European Union average	404.105	9.469

Table 8 shows that Norway invests more in R&D than the European Union average. The first column shows gross domestic expenditure on R&D per capita, while the second shows total R&D personnel in relation to total employment (in thousands).



## Appendix B

To show that a Dixit-Stiglitz consumption index implies a constant elasticity of substitution between any two products larger than one, suppose that the utility function is the Dixit-Stiglitz index:

$$D = \left[ \int_0^n x(j)^\alpha dj \right]^{1/\alpha}, \quad 0 < \alpha < 1$$

To simplify the maximization problem, replace  $D(t)$  with  $D(t)^\alpha$ , which is a strictly increasing transformation of  $D(t)$ . Then use Lagrange's method to maximize the consumption index subject to the budget constraint  $E = \int_0^n p(j)x(j)dj$ . The Lagrangian becomes:

$$L = D^\alpha - \lambda \left( \int_0^n p(j)x(j)dj - E \right)$$

The first derivative of this is:

$$\frac{\partial L}{\partial j} = \alpha x(j)^{\alpha-1} - \lambda p(j) = 0$$

This holds for all varieties  $j$ . Solving this for two varieties  $j$  and  $j'$  yields:

$$\frac{x(j)}{x(j')} = \left[ \frac{p(j)}{p(j')} \right]^{\frac{1}{\alpha-1}}$$

This shows that the elasticity of substitution is  $\varepsilon = 1/(1 - \alpha) > 1$ .

## Appendix C

Table 9 contains the sector names for all manufacturing sector codes.

Table 9: NACE Rev. 1.1

Sector code	Sector name
15	Manufacture of food products and beverages
16	Manufacture of tobacco products
17	Manufacture of textiles
18	Manufacture of wearing apparel, dressing and dyeing of fur
19	Tanning and dressing of leather, manufacture of luggage, handbags, saddlery, harness and footwear
20	Manufacture of wood and of products of wood and cork, except furniture, manufacture of articles of straw and plaiting materials
21	Manufacture of pulp, paper and paper products
22	Publishing, printing and reproduction of recorded media
23	Manufacture of coke, refined petroleum products and nuclear fuel
24	Manufacture of chemicals and chemical products
25	Manufacture of rubber and plastic products
26	Manufacture of other non-metallic mineral products
27	Manufacture of basic metals
28	Manufacture of fabricated metal products, except machinery and equipment
29	Manufacture of machinery and equipment n.e.c.
30	Manufacture of office machinery and computers
31	Manufacture of electrical machinery and apparatus n.e.c.
32	Manufacture of radio, television and communication equipment and apparatus
33	Manufacture of medical, precision and optical instruments, watches and clocks
34	Manufacture of motor vehicles, trailers and semi-trailers
35	Manufacture of other transport equipment
36	Manufacture of furniture, manufacturing n.e.c.
37	Recycling

## Appendix D

The input-output table is taken from Statistics Norway. It shows the trade in intermediate inputs among the manufacturing sectors in Norway in 2001. The rows indicate how much the sector is producing and selling to other sectors, while the columns show how much the sector is purchasing from other sectors. The numbers along the diagonal are the intra-industry trade in intermediate inputs.

Table 10: Input-output table (2001)

Sector	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
15	30502	10	32	9	70	94	69	177	12	768	30	172	153	159	201	6	98	107	55	25	327	89	18
16	29	1	2	0	0	6	4	5	3	13	3	6	13	7	14	1	5	5	5	2	17	4	1
17	128	3	609	266	12	16	46	76	2	43	120	30	40	26	43	1	46	10	72	15	153	334	3
18	47	1	3	15	1	13	7	17	1	25	6	14	39	23	34	1	8	12	8	8	36	8	2
19	12	0	1	3	72	4	2	8	0	6	2	3	11	8	10	0	2	3	3	3	9	238	1
20	168	3	6	2	1	3503	155	32	3	93	8	45	37	69	75	1	17	13	9	18	291	1103	3
21	1251	31	23	5	3	125	3516	3651	16	293	87	218	117	58	76	0	51	43	17	6	38	77	6
22	415	5	20	7	3	72	73	4810	12	216	38	115	117	114	172	4	67	74	95	30	224	68	12
23	268	4	11	3	2	54	217	25	1379	772	18	337	733	40	43	0	365	2	4	11	80	51	17
24	1749	5	379	5	19	436	697	712	109	8382	1604	389	889	513	382	3	583	110	84	130	821	396	13
25	1354	26	34	5	3	134	106	124	16	278	407	98	97	163	309	9	192	297	38	133	170	335	70
26	300	5	9	3	1	365	30	42	4	156	113	1416	331	364	141	2	213	111	90	17	295	74	5
27	121	2	60	2	1	126	33	81	2	204	41	137	14509	3580	2922	1	1471	138	111	992	3426	594	660
28	712	13	75	5	6	297	76	88	7	281	70	172	309	1318	1345	4	413	112	152	206	4422	214	15
29	632	8	24	6	4	124	197	221	5	475	62	144	425	390	6800	5	188	76	57	265	3698	85	43
30	114	1	3	1	1	24	46	40	0	73	12	24	70	20	233	535	59	489	733	8	50	12	10
31	213	2	8	2	1	45	67	85	2	157	20	54	662	128	759	5	1007	227	453	53	1894	45	16
32	80	1	4	1	1	15	19	32	1	60	6	20	261	56	424	105	277	2984	1045	5	128	10	4
33	63	1	3	1	0	12	15	24	1	37	5	15	29	18	63	1	15	16	856	4	742	8	3
34	30	1	3	1	1	7	5	30	1	22	2	8	63	9	161	0	29	9	5	587	87	6	1
35	616	7	23	6	3	75	56	34	2	302	19	176	120	178	373	4	97	89	37	53	9147	60	18
36	100	1	23	7	1	59	16	41	3	51	8	29	27	27	41	1	15	15	11	6	591	713	3
37	27	0	1	0	0	4	845	14	1	12	39	5	810	200	9	0	3	4	3	2	11	4	1